経済学研究 53−3
北海道大学 2003.12

倫理的規範形成のゲーム論的分析

町 野 和 夫

1．はじめに

個人的自由が広く認められている大きな社会が破綻せずに存続するためには，社会的ジレンマやフリーライダー問題が解決されなければならない。経済学でも，公共財の供給，外部性，非対称情報など「市場の失敗」が存在する場合は，個々の経済主体が自由に行動する完全競争市場でパレート効率的な結果が得られないことは，基本的知識となっている。従来の経済学ではそのような「市場の失敗」を，政府の介入，保険市場の創設，当事者間交渉の場の設定などによっていかに補正するか，という方向で議論されてきた。その場合経済学が想定する行動主体は，コスト・ベネフィットを計算して目的関数（消費者であれば効用関数，企業であれば利潤関数）を最大化する合理的プレーヤであった。

これに対して，近年，囚人のジレンマや交渉ゲームなどの実験結果と理論モデルが予測する結果とのギャップを説明するためには，利他的もしくは互恵的効用を持つプレーヤの存在を仮定すべきだという考えが定着しつつある。これはミクロ経済学理論の見直しの動きとも合致している。即ち，ゲーム理論モデルで複数均衡を絞り込むための均衡精緻化が進んだ結果，あまりにも合理的になった均衡概念の非現実性への反省という動きである。

経済理論における主体の過度の合理性に対する批判は，既にサイモンの限定合理性の議論（Simon, 1956）にも見られたが，最近の限定合理性の議論は，1970 年代頃から生物学で使われ始めた進化ゲームの枠組みや，認知心理学で蓄積されてきた様々な学習モデルを応用して，慣習や規範のモデル化で具体的な成果を上げている（例えば Fudenberg and Levine, 1998）。

主流派経済学は伝統的に社会的価値判断を分析の対象外に置いてきた。市場での自由な競争の結果もたらされるパレート均衡は複数存在しうる。それらの間の比較を試みる厚生経済学，社会選択論という分野も存在するが，それは「何が倫理的か」を論じているのであって（鈴村 2001），本稿で問題にしている，「どのように倫理規範が形成されるか」を対象としている訳ではない。

進化ゲームや学習ゲームを使って慣習や規範の形成を説明しようとするモデル（Kandori *et al.*, 1993, Young, 1993 など）の中で，例えば Young は，記憶に限界のある限定合理的なプレーヤが，記憶に残っている自己と相手の戦略の組合せと結果（＝利得）に基いて，最適戦略をプレーするモデルを考えた。確率過程モデルなので同一戦略の組合せが何回か連続して起こる可能性が存在し，さらにプレーヤの記憶に限界があることから，同一の戦略と結果が，それのみが記憶に残る程連続的に生じる可能性がある。するとその間最適戦略も同じなので，お互いの取る最適戦略が長期にわたって同じもの，即ち慣習となる。さらに長期的に見ると，利得の高い戦略がとられる頻度が高いので，慣習は合理的なプレーヤによる均衡と同じである可能性が高い。この種のモデルが倫理的規範形成について言えることは，もし倫理的行動（という戦略）が高い利得をもたらす（最適戦略）なら，それが規範として慣習化する可能性があるとい

うことである。

倫理的行動は自己犠牲的な行動パターンである。確かにフォーク定理が示すように，長期的な最適反応としてプレーヤ間の協調が均衡となる場合があり，それが慣習化する場合もあるであろうが，実験経済学の結果（例えば Fehr and Gachter, 2000）が示しているのは，（上述の規範形成の進化ゲーム・モデルでも前提としていた）長期的な合理性を超えた「自己犠牲」を厭わない倫理的行動の存在である。もしこのような行動が存在することを前提にすると，それは長期的に安定的な戦略となる可能性をモデル化したのが，Bowls and Gintis（2003）である。彼らは利他的行動，互恵的行動，集団主義的行動などの戦略をモデル化し，同様のタイプのプレーヤがある程度の比率で存在する限り，そのグループとしての存続の力が強まることを証明した。また，Axelrod（1984）に始まるコンピュータ・シミュレーションに基づく一連の研究も，社会における利他的，互恵的プレーヤ・グループの存続可能性を強く示唆している（例えば，Axtel *et al*., 2001）。

彼らとは異なるアプローチとして，行為の「倫理性」そのものから受ける正の効用を説明することを重視したのが，心理ゲームである。心理ゲームでは，相手の戦略集合と利得関数，実際に相手が選んだ戦略から，相手の行動の「意図」を推測し，相手の意図が好意的かどうかによって，自分の戦略が相手の利得に与える結果に対する自分の利得が左右される（Falk and Fishbacher, 2000）。換言すれば，好意的な相手だと判断すれば，自分の戦略集合の中で相手の利得を上げる戦略を選択すれば，自分の効用も上がり，相手の利得を下げる戦略を選択すると自分の効用も下がる。逆に相手が好意的でないと判断すれば，効用の方向も反対になる。

しかし，互恵的，利他的，集団志向的戦略の進化的安定性の条件を明らかにした前者のモデルは，倫理規範が成立した後の安定性の条件を求めたものであり，互恵的効用を組み込んだ後者の心理ゲームも倫理規範が既に内面化されているプレーヤを前提としており，両者とも倫理的規範の形成を説明したモデルではない[1)]。

2．本稿のアプローチ

本稿では，次の二段階に分けて倫理規範形成過程モデルの概要を示す[2)]。第一段階では，実際に個々の人間がどのようにして倫理規範を身に付けてきたかに着目し，そのモデル化を試みる。倫理観の形成，即ちある行為自体を高い価値を持つ「善いことである」と信じさせるのは，子供の頃の家庭教育や学校教育などの社会化の役割である。もちろん，その後，所属する社会集団の影響や個人的経験・思索を経て個人の倫理観は変化するが，それが可能になるのは，幼少期に意思決定のための価値判断の尺度として，自分の欲望以外に善悪の概念（倫理規範）を身につけているからであろう。

倫理規範を教える側の動機について考えると，利己的なものもあるが倫理的なものもある。教える側の動機が倫理的なものだとすると，それを教えた者の動機は何か，そしてそれを教えた者の動機は... と，因果関係に関する議論は無限に遡る可能性がある。従って，倫理規範形成過程モデル化の第二段階として，そうした無限の連鎖を断ち，倫理規範の合理的形成の論理を提供しなければならない。この段階で重要なのは，何らかの自己犠牲的行動を他の共同体メンバーに説得する，提案者（リーダー）の存在である。説得された行為が共同体にとって有益であることが実証されれば，この行為が次第に「倫理的行為」として認識され，その認識が共同体に浸透していく。

以上のような倫理規範形成の理解を前提とし

1）その他，規範形成のモデル化という意味で前者に属する議論としては，Kaneko and Matsui (1997)，Aoki（2003），Kandori（2003）などの議論がある。

2）個々のモデルの詳細は町野（2003）を参照。

て，本稿では倫理規範の形成過程を，社会のルールとしての倫理規範を大人が子供に躾ける社会化過程と，ある行為が社会のルールであるべきだと提案者が社会（共同体）に説得する過程に分けてモデル化する。前者では，とくに大人のインセンティブを厳密に考えることで，従来の学習モデルとは異なるゲーム構造を明らかにする。後者では提案者（リーダー）のイニシアティブの重要性を明らかにし，従来の動学的進化ゲームに欠けていた倫理性発生のモデル化を試みる。これらの過程と，慣習・規範化に関する近年の研究成果を統合することで，倫理規範形成過程の体系的モデルの概要を提示する。

3．社会化

3-1　考え方

プレーヤは規範を教える大人（家族，地域社会，学校，職場，宗教組織など）と教えられる子供である。子供は何も倫理規範を身につけていない状況から，社会化の過程で倫理規範が内面化される[3]。

子供が倫理的行動を習得する過程をモデル化するには，限定合理性を前提とした学習ゲーム，例えば最適反応などのモデルを使えばよい[4]。これは，叱られる不効用がその行為から得られる効用を上回るときに，子供が叱られないようにするにはどうすればいいかを徐々に学んでいくメカニズムである。しかし，これらのゲームにおける子供は，利己的な動機から倫理的な行動を選択することが得であるということを学ぶだけで，「倫理的になる」わけではない。

初めは倫理的でない子供が倫理的になる状況をゲーム・モデルで表現するには，当初は負の効用をもたらす（自己犠牲を伴う）倫理的行動が，最終的には正の効用をもたらすようになる過程が説明できなくてはならない。そのような過程の例として，条件反射という学習過程がある。条件反射の類推で考えると，この過程は，子供が反倫理的行動を取ったときに，常にその正の効用を打ち消す制裁（叱る，もしくは諭す）を与えて，次第に反倫理的行動自体が負の効用をもたらすようにする，言わば躾の過程である。

倫理的行動（あるいは反倫理的行動）が何かは分かっていると仮定すると，この過程では，子供の「反倫理的行動」と，大人の「叱る」という行動の組である戦略プロファイルが実行される度に，子供は常に負の効用を得るため，その戦略プロファイルと負の効用が対になって記憶として蓄積される。さらに，人間の記憶が限定的である（あるいは理解力不足の）ため，マイナスの効用が相手の意図的行動の結果であるという部分の記憶が省略され，反倫理的行動をとると負の効用を得るという，一段階節約された効率的反応メカニズムが形成される。この節約されたメカニズム，即ち反倫理的行動をとることによってもたらす負の効用を，倫理感と呼べるだろう[5]。このような躾の結果，「悪いこと」をすると罪悪感を抱くようになる。逆に「倫理的行動」と「褒める」という戦略プロファイルであれば，正の効用と結びつき，（自己犠牲を払っても）善いことをすると満足感を得られるようになる。

もちろん倫理感という感情が成熟するには，自分の行為に対する他者からの評価及びそれに伴う正もしくは負の心理的刺激（賞賛や叱責）の記憶が，自分の行為と心理的刺激を直結させるまで繰り返されなければならない。他人との関係で言えば，自分の行為と他者の反応の因果関係を（おそらく無意識のうちに）納得して，

3）社会化，子供の心理的発達については，例えば斎藤・菊池（1990），無藤他（1995）を参照。
4）例えば前掲，Fudenberg and Levine 参照。
5）この倫理感形成過程は，制裁の繰り返しによってのみ実現するとは限らず，信頼する他者の模倣，説得と納得という過程を経て実現するかもしれない。

自分の行為に対して他者が与える制裁や賞賛を「当然のもの」と考えるようになるまで何度も経験しなければならない。

社会学で内面化（あるいは同一化）と呼ばれる過程がこれにあたる。「内面化」の社会学的説明としては，例えば作田（1972）は，自分が全面的に依存している大人との葛藤を避けるために，大人の価値観を自分のものとする過程と説明している[6]。内面化されるには，他者による社会的評価（及びそれに伴う反応）の記憶がある臨界値を超えて蓄積されなければならない[7]。もちろん一回ごとの刺激の強さ，次の刺激までの間隔の長さ，類似行為が内面化されているかどうかによって，臨界値に達するのに必要な繰り返しの回数は異なる。

以上の議論では大人側の動機には触れてこなかった。倫理規範を教える大人の動機には様々なものが考えられる。（家族のメンバーである）子供がその社会で生き残るための掟を身につけることを望む本能的な愛情，子供が社会から認められることの誇りとその逆の場合の恥，その集団全体の利益のため，若いメンバーの身勝手な行動をコントロールすること，自分の監督下にあるメンバーが，他のメンバーもしくは外の社会とうまくいかなかったときに自分が被る費用を最小化すること，単に自分の好む行動を取ったときの正の効用，逆に自分の嫌う行動をとった場合の負の効用，などである[8]。

叱ったり褒めたりする教育的行為が，大人本人の直接的費用・便益分析から合理的に導かれる場合は，その社会に，子供が倫理的行動をとらないことの負の外部効果を大人に負担させるメカニズムが既に組み込まれていることになる。具体的には，非倫理的行動のモニタリングや，子供の行為に対する大人への費用負担強制のメカニズムが法制化され，運用面でも現実的なコストで機能しているということである。このように，大人の利己的動機によって子供に倫理的規範を教える直接的メカニズムも重要な問題ではあるが，本稿の主旨からは外れるので，この点はこれ以上立入らない。

利己的動機ではなく，叱ったり褒めたりする教育的行為を，倫理的責任あるいは義務と感じている場合には，教育行為という費用を払ってまでも心理的満足感を得ること（としなかった場合の罪悪感を避けること）が倫理規範を教える動機になる。しかし，教育的行為自体から効用を得るのはなぜだろうか。

いま仮に倫理的行動を，その状況で当然取るべき行為と仮定しよう。したがって，子供が倫理的行動をとらないと，その状況でとられるべき行動が取られなかったので，大人は期待を裏切られる。また，その結果生じる実際の効用は元々の期待効用に比べて低い。原因は他者（子供）の行動と自分の期待が異なったためだが，限定合理的プレーヤである大人は，子供の行為自体が負の効用をもたらすと認識する。すると，このような負の効用をもたらす反倫理的行動を阻止する手段があれば，それは効用を増加させる行動ということになる。例えば反倫理的行為をとったプレーヤ（子供）を叱ることで，相手がその行動を止めるという結果が繰り返されると，制裁行動自体が無意識に正の効用と結び付けられていく。ただし，その制裁行動の費用が，相手に倫理的行動を取らせるときの効用増（期待値）を上回れば，長期的には維持不可能である。（短期的には，怒りや防衛本能と結びつけば，能力の範囲で最大限の制裁行動に出る可能性もある。）

ここで最後に残る疑問は，今仮定した，「その状況で当然取るべき行為」としての倫理的行為はどのように決められるのかということである。これは前項で述べた共同体の倫理規範形成

6）作田（1972）pp. 95-124。
7）倫理観を持つということと，倫理的行為を行うということは同じではない。倫理的行為の実行に，正の効用を上回る費用がかかれば，その行為は実行されない。この点は後述する。
8）その他，生物学的な親の本能という要因もある。（例えば前掲，無藤他第 10 章参照）

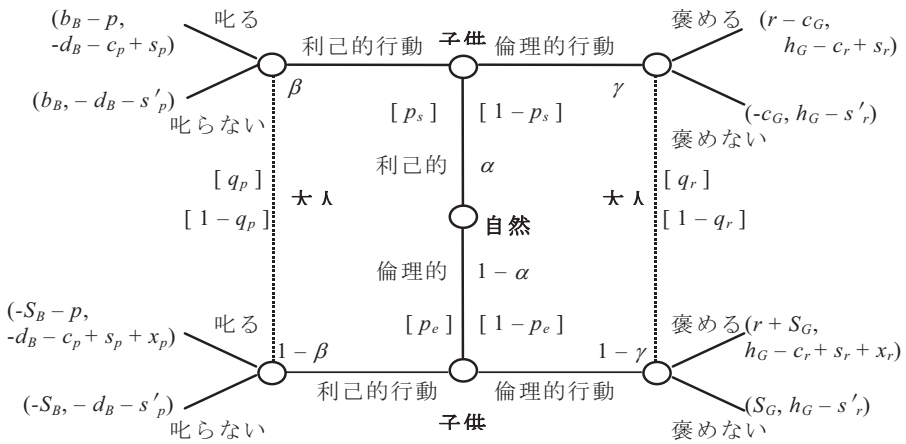

図１　社会化（大人の視点）の構成ゲーム

問題であり，具体的メカニズムは次節で扱う。

3-2　モデルの概要

上述のように，この過程は条件反射的学習過程である。子供は倫理的行動が何であるかを知らない，つまり大人の戦略集合，効用関数という情報を持たないが，大人は，子供の戦略集合や効用関数の情報を全てあるいは一部持っている。教育過程としては，大人がまず先に動く（教える）という場合も多いが，ここでは倫理的行動が何であるかを知らない子供が，叱られたり褒められたりしながら倫理感を身につけていく過程を想定しているので，子供が先に動き，大人はそれに反応して後から動く。上述のように，本節の課題は子供が倫理規範を身につける過程のモデル化であり，大人は既に倫理的規範を反映した利得，戦略集合を持つと仮定している。大人の倫理的規範の形成は次節で明らかにする。

通常のベイジアン・ゲームと異なり，子供は当初大人の戦略集合と効用関数についての情報を全く持っていない。両者の知識構造の違いを明確にするため，両者の異なる視点で，ある時点での構成ゲームの木を描くと図１と図２になる。まず大人の視点で描かれるゲームの木は図１であり，その流れは次のとおり。

① 自然が子供の利己的である確率（α）を決める。(α は当面所与の定数)

② 子供（C）が反倫理的行動（B）か倫理的行動（G）を選ぶ。（実際は初めから倫理的行動を選ぶことはない。）[9] 利己的な子供と倫理的な子供の行動戦略をそれぞれ（p_s，$1 - p_s$），（p_e，$1 - p_e$）とする。

③ 大人（教育者）（T）は子供の行動が反倫理的行動であれば失望し（$-d_B$），倫理的行動であれば嬉しい（h_G）。彼らはまた子供の行動が反倫理的行動であれば叱るという制裁（*pn: penalty*）を与え，倫理的行動であれば褒めるという報酬（*rw: reward*）を与える。そのような教育的行動には心理的，時間的費用（叱る，褒める，それぞれ c_p と c_r）と責任を果たしたという心理的満足感（叱る，褒める，それぞれ s_p と s_r）が伴う。逆に，教育的行動をとらない（*nrw* あるいは *npn*）と後悔（叱らない，褒めない，それぞれ$-s'_p$

9) 現実には利己的行動でも反倫理的とは限らず，叱る必要の無い行動も多い。しかし，簡単化のためここではそのような中立的な行動は考えない。中立的行動を明示したモデルでも同様の結果が得られる。

と$-s'_r$）する。子供が倫理的であれば，教育的行動に反応するのでさらにプラスの効用を得る。（叱るとき，褒めるとき，それぞれx_pとx_r）

④　子供の教育的行動への反応を見て，子供の利己的な確率に対する信念を修正する。

ただし，大人の認識する利己的子供の利得は，反倫理的行動からの効用がb_B，倫理的行動の効用が$-c_G$，倫理的子供の倫理的行動からの効用がS_G，反倫理的行動の効用が$-S_B$であり，叱られたときと褒められたときの効用はそれぞれ$-p$，rである。

大人は，子供の行動が反倫理的行動であれば失望し（$-d_B$），倫理的行動であれば嬉しい（h_G）。このことが前提に無ければ次の教育的行動は誘発されないが，それだけでは，大人が褒めたり叱ったりする直接の条件にならない。実際に褒めたり叱ったりするのは，そのような教育的行動を取ったときの利得が，教育的行動を取らなかったときの利得を上回るからである。前者は責任を果たしたという心理的満足感（s_pまたはs_r）から，教育的行動にかかる心理的，時間的費用（c_pまたはc_r）を差し引いた利得，後者は，教育的行動をとらないときに後悔する心理的費用（s'_pまたはs'_r）である。

このゲームは典型的なシグナリング・ゲームであり，パラメータの値，具体的には両プレーヤの利得構造とαの値によって，利己的な子供であっても倫理的な行動を選択する均衡が存在する。

次に，躾けられるに従って，反倫理的行動からの子供の効用が，b_Bから$-S_B$に，倫理的行動からの子供の効用が$-c_G$からS_Gに変化する（利己的な子供が倫理的になっていく）仕組みは，以下のような子供の立場からの二段階のゲームを考えることで説明される。

第一段階として，子供の視点で描かれる構成ゲームの縮約された木は図２であり，その流れは次のとおり。

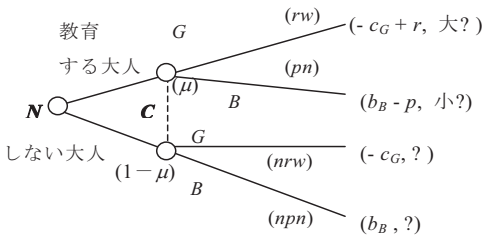

図２　社会化（子供の視点）の構成ゲーム

①　子供は当初，大人が自分の行動を褒めたり叱ったりするとは予想しない。即ち図２のゲームの木を無意識に認識することもなく利己的に行動する。（$\mu=0$）

②　子供は利己的に行動をして，それが反倫理的行動（B）であれば，利得（b_B）は得られるが，叱られ（$-p$），倫理的行動（G）で犠牲（c_G）を払うと褒められる（r）ことによって，大人に叱られる「悪いこと」と褒められる「善いこと」があることを認識し始める（$\mu>0$）。子供は叱られたことを忘れ易く，本能的な欲望を抑えにくいので，信念を正しい方向に修正して，図２の構造を（直感的に）理解するのに時間がかかり，何回かこのゲームを繰り返さなければならない。

次に第二段階では，このような倫理規範教育のゲームが構成ゲームとして行われ，それが繰り返されることによって，子供にとって，当初利得が高いという利己的動機から行っていた「倫理的行動」が，徐々にその行為自体から利得を得るようになる。これが，子供の「社会化」，即ち倫理規範の形成である。その過程をモデル化するとおよそ次のようになる[10]。

子供（C）のt期の戦略を，$a_t\in\{G, B\}$（G, Bの定義は上記モデルと同じ），子供のt期の戦略がa_tであったときの大人（T）の戦

10）モデルの詳細は町野（2003）を参照。

略を，$z_t = z(a_t)$，t 期までの（戦略プロファイルの）歴史を k_t をとする。ここで，単なる繰り返しゲームと異なり，子供には叱られたり褒められたりする記憶が（時間が経つに連れて薄れるものの）徐々に定着するという躾のプロセスを考える。例えば過去に（B, pn）という戦略プロファイルを n 期続けて繰り返すと，その記憶は $n+1$ 期の初めには，$0 < \rho < 1$ という記憶の不完全さを示すパラメータを使い，$(\rho+\ldots+\rho^n)(b_B-p)$ として残る。

躾けられていない子供は限定合理的なプレーヤで，上述の躾による限定的な記憶とその時点で取る行為からの直接的な効用だけが期待効用の要素であるとする。したがって，0 期にはまず反倫理的行動を取るとすると，その後大人の教育的行動が始まるが，$a_t = G$ の期待効用が $a_t = B$ の期待効用を上回るまでは，（B, pn）という戦略プロファイルを取る構成ゲームが繰り返され，ある時点 s 期に，子供の戦略が B から G に変わる均衡経路が存在する。

子供の戦略が a_t，歴史が $k_t = \{(a_1, z_1)_1, \ldots, (a_{t-1}, z_{t-1})_{t-1}\}$ であると，t 期期初の子供の期待利得は，この均衡経路上では s 期の前後に以下のように変化する。

(i) $t < s$ のとき，

$$\mathrm{E}U^C{}_t(B;\ k_t{}^*) = b_B+(\rho+\ldots+\rho^{t-1})(b_B-p) > U^C{}_t(G,\ k_t{}^*) = -c_G$$

ただし，

$$k_t{}^* = \{(B,\ pn)_1, (B,\ pn)_2, \ldots, (B,\ pn)_{t-1}\}$$

(ii) $t = s$ のとき，

$$\mathrm{E}U^C{}_t(B;\ k_t{}^*) = b_B+(\rho+\ldots+\rho^{t-1})(b_B-p) < U^C{}_t(G,\ k_t{}^*) = -c_G$$

$$k_t{}^* = \{(B,\ pn)_1, (B,\ pn)_2, \ldots, (B,\ pn)_{t-1}\}$$

(iii) $t > s$ のとき，

$$\mathrm{E}U^C{}_t(G;\ k_t{}^*) = -c_G+(\rho+\ldots+\rho^{t-s})\times(r-c_G) > U^C{}_t(B,\ k_t{}^*) = b_B+\rho^{t-s}(\rho+\ldots+\rho^{s-1})(b_B-p)$$

ただし，$k_t{}^* = \{(B,\ pn)_1, (B,\ pn)_2, \ldots, (B,\ pn)_{s-1}, (G,\ rw)_s, \ldots, (G,\ rw)_{t-1}\}$

(i)～(iii)が成り立つのは各期待効用が t の単調増加関数もしくは単調減少関数であるからだが，注意すべきは，s 期以後の $\mathrm{E}U^C{}_t(B;\ k_t{}^*)$ が単調減少することである。即ち叱られていた記憶が薄れて反倫理的行動の期待効用が倫理的行動の期待効用を再び上回る可能性がある。ここではパラメータがそのような可能性を排除する値であったとすると（ρ が十分大きければよい），このプロセスは s 期以後に，行為→結果→効用が繰り返されることによって，情報節約のため行為→期待効用という流れに短縮化されたものであり，条件反射と同様の構造であると解釈できる。

4．倫理行動の規定とリーダーのイニシアティブ

4-1　考え方

前項では，大人が利己的，倫理的いずれの動機から躾を行うにせよ，大人の側は既に何らかの倫理規範を身につけていることが前提となっていた。倫理規範形成を考えるに際して，大人の倫理規範がどう形成されたかを説明する（躾とは別の）メカニズムが無ければ，社会化の議論はその大人が子供のときの大人，その前の世代の大人，という具合に無限の後退を続けるだけである。

しかし無限の後退も一面の真理を衝いている。幼児期以後の共同体や社会における「学習」によって，幼児期の倫理観は修正され体系化されていく。子供が初めて自分の倫理的基準に基づいて行為を判断し始めるようになった時，その行為は心理的満足感を伴うとしても，少なくとも何らかの「自己犠牲」を伴う行為である。最初は，大人（養育者）が高い価値を付加するような自己犠牲行為を倫理的だと判断するであろうが，その基準はまだ感覚的であやふやなので，様々な場面で大人から褒められたり叱られたりする経験と，遊び仲間など大人以外から新たなルールを学ぶことによって，変化しつつ固まっていく。大人にしろ遊び仲間にしろ，彼らが身につけている価値観は，長い年月をかけてその

地域で受け継がれてきた価値観を反映しているので，幼少期に無意識に身につける価値基準は，平均的には共同体あるいは社会の価値観と整合的であると仮定できる[11]。

では，共同体や社会の倫理規範はどのようにして形成されたのであろうか。倫理という概念の有無に拘らず，個々の人間が共同体の中で生きている以上，その個人の行為は，他の共同体メンバーの効用にとって，中立であるか，そうでなければ正か負の外部性を与える。一般にはある行為をめぐる共同体メンバーの利害は対立関係を含むが，自分も含め共同体全体の便益が増加する公共財のようなケースもある。全ての人の効用が増加するパレート改善が期待できる場合でも，囚人のジレンマ（あるいは社会的ジレンマ）のように他人の行動に対する信頼が持てなければ，自分にとっても共同体全体にとっても，より低い効用をもたらす行動をとってしまうこともある。パレート改善を実現する行動を選択させるには，その行為をしない場合のコストを高める（たとえば強制する）か，それが結局はその人にも得になることだと説得するしかない。強制にしても説得にしてもコストがかかる。いまは倫理的規範が無い世界を考えているので，強制や説得のコストを上回る利益を得られなければ，その行為のイニシアティブをとる人はいない[12]。

仮に共同体のあるメンバーが，他のメンバーに対して，ある協同作業が公共的な利益になると提案したとしよう[13]。 倫理的動機は無いと仮定しているので，この協同作業の提案は，説得する費用に比べて，提案者にとって私的及び公共的な期待利益の合計が大きいはずである。しかし提案者の利害と説得される側の利害が一致するとは限らない。仮に提案者の説明が正直なものであっても，効果に対する見通しが誤っている場合もある。また，逆に説明が不正直なものであっても，結果的に正の外部性が発生することもあるので，提案者の利益の公共的な部分を事後的にも正確に知ることは出来ない。つまり，提案者が自身にとって利益になると思っていることは確かだが，説得される側にとって得かどうかは分からない。このような状況で共同体メンバーが説得されるときの説得力の源泉は何だろうか。

無意識に倫理を内面化させる場合と違って，この例では論理的或いは感覚的に「納得」しなければ，提案は受け入れられない[14]。共同体のメンバーが，「自己犠牲」という費用を求める提案者の正当化の論理に納得するのは，最終的には自分にメリットがあると信じられるときである。少なくとも長期的には自分のためになるという理屈を，多くの共同体メンバーが「正しい」と思って初めて，提案が共同体に受け入れられる。従って，この説得で使われる理屈は，共同体でその正当性が共有されている因果関係でなければならない。

しかし当初は，提案者が正直に説明しているかどうかという非対称情報の状況で，しかも提案者の見通しが正しいかどうかという不確実性がある中で，少なくとも提案された行為を実行してもらえる水準まで納得してもらわなければならない。事前に何の先入観も無いとすれば，提案者が正直である事前確率は二分の一である。

11）大人と共同体の倫理基準が矛盾する場合は，大きな葛藤が避けられない。また第二次世界大戦前後の日本のように，社会の価値観に大きな断絶がある場合もあるが，そのようなケースはここでは考えない。

12）複数の人が同時に「自己犠牲的」行為のメリットに気づけば，より少ないコストで倫理的行為が広がることもあるが，そのような幸運なケースはここでは扱わない。

13）共同体は成立していても，共同体のための行為は，強制や慣習によって行われると仮定すれば，倫理的動機が存在しないと仮定できる。

14）説得する人が既にリーダーという地位に就いているなら，彼もしくは彼女がこれまでの実績の積み重ねによって，共同体のメンバーから信用されているかもしれないが，後述するように，本稿ではその点は捨象してモデル化する。

それを少しでも高めるための説得が必要である。また，説得する側の提示した因果関係，具体的にはその協同行為が成功する確率や利得の情報についても納得してもらわなければならない。通常のモデルではこのような情報については共有知識と仮定されるが，それらが共有知識になるためには，提案者の使う論理が，説得される側の信じている因果関係と整合的でなければならない。

現実には共同体が成立していることを前提にすると，共同体メンバーの間で，それぞれのメンバーが正直である事前確率は，ある程度メンバー間に共有されていると思われる。すると各メンバーの提案者としての信頼性の高さは，共同体の中でほぼコンセンサスがあることになる。また，提案者の使う論理も，他のメンバーが評価できないほど独創的な場合を除けば，共同体のメンバーが共有している因果関係から大きく外れるものは少ないと考えてよいだろう。即ち，提案者の信頼性，彼（女）の提案する協同行為の主観的成功確率，費用便益効果の期待値は，概ね共有知識であり，違いはメンバー個々の効用関数の違いと見なすことができるであろう。したがって，提案者が正直である事前確率も，その協同行為が成功する事前確率や期待利得も，説得が始まる前に共同体の歴史の中で決められている。説得過程全体にとっての問題は，どの程度の信頼性を持つ提案者が，数ある協同行為の可能性の中で，どのくらいの成功する事前確率や期待利得を持つアイディアを思いつくか，ということになる。

最初に説得を始めるのが誰かということも，他のメンバーの反応を大きく左右する。提案するのが他のメンバーから既に信頼を得ている者であれば，提案の説得力は高くなる。それが実績の無い新しいメンバーであれば，他のメンバーを説得するのは難しい。従って，生得的要素である体力，知性，経験など，何らかの点で他のメンバーに勝っているメンバーが次第にリーダーとなって，自らの利益のため，共同体の利益にもなる協同作業を提案すると考えた方が自然かもしれない。しかし本節では，まずリーダーとしての影響力を前提せずに，提案者はある協同作業を思いついた点だけが，他のメンバーと異なるという設定で考える。

一旦説得を受け入れて協同作業を行った後は，結果として公共的利益が費用を上回るほど高ければ，その協同作業の公共性に対する他のメンバーの信頼が高まってくる。不確実性もあるので，常に共同体にとっての利益を実現できるとは限らないが，長期的に公共的な利益を増進できれば信頼されるようになる。つまり，歴史的な実績の積み重ねがあれば，当初の意図に拘らず信頼は醸成される。

この過程も第一段階はベイジアン・ゲームとしてモデル化できる。現実の結果を見て個人の自己犠牲の費用を上回る利益が実現すれば，その行為が「善いこと」である事後確率をより高い数値へ修正するし，逆の場合はより低い数値へ修正する。

以上の状況を素直にゲーム・モデル化すると，提案者が一人，その他の共同体メンバーが多数のゲームになる。本稿ではその複雑さを軽減する工夫として，説得されるかどうかは提案者と個々のメンバーの二人ゲームとし，説得される人数がある閾値を越えるかどうかに利得を依存させるゲームと考える。閾値を越えれば，費用を上回る公共的便益が実現される。

4－2　モデルの概要

提案者と個々のメンバーによる次のような流れと利得を持つ二人ゲーム（図３）が，構成ゲームとして長期的に繰り返される。以下の過程が，まず言葉による説得によって行われ，そこで納得されれば，以後は実績による再評価も加わって行われる。最初の言葉による説得の際には，共同体メンバーは，利得関数のパラメータ，提案の公共性についての事前確率，協同作業の成功確率を，過去の類似の経験や提案者の説得を勘案しながら設定し，受容したときと拒否した

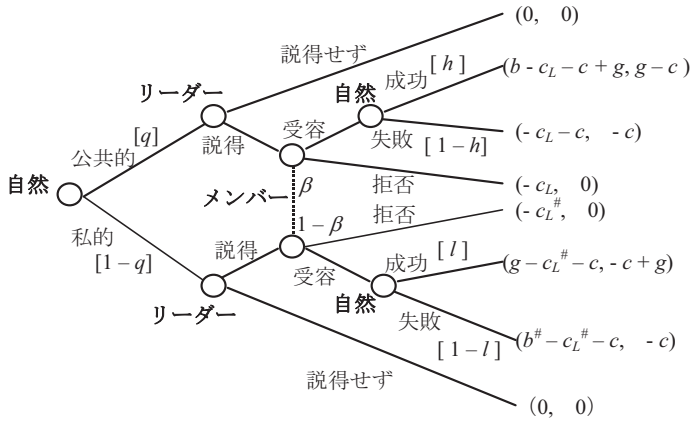

図３　説得ゲームの構成ゲーム

ときの期待利得を比較して，諾否を決定する。そこで受容されれば，実際に④〜⑥のステップが実行される。

① 自然（N）が，ある協同作業のタイプを，公共的利益の実現可能性の高いタイプと，主に提案者の私的利益を実現するタイプの二つから一つ選ぶ。

② 提案者は，いずれのタイプの協同作業であっても，自己の期待利得が正で大きい方の行動を取るように他の共同体メンバーを説得する。このときの説得コストは，公共的提案の方（$-c_L$）が，私的提案の方（$-c_L^{\sharp}$）より低い。

③ 説得を受けた共同体メンバーは，説得を聞き入れるか拒否するか決める。拒否した場合ゲームは終わり，提案者には説得費用がかかり，両者とも便益は無い。

④ 提案者の説得を聞き入れたメンバーは自己犠牲（$-c$）を伴う行動をとる。提案者自身も同じ行動をとる。

⑤ 説得された協同作業が公共的利益を実現するかどうかは，自然（N）が決定する。ただし公共的利益実現の可能性は，公共的提案の方（h）が，私的提案（l）より高い。

⑥ 公共的利益が実現した場合は，両プレーヤとも自己犠牲的行動のコストを上回る公共的利益（g）を得るが，このとき公共的提案であれば，提案者はさらに私的利益（b）を得，私的提案であれば，提案者には追加的私的利益はない。逆に公共的利益が実現しないときは，公共的提案であれば，提案者も追加的利益を得られないが，私的提案であれば提案者は私的利益（$b^{\sharp}$）を得る。

このゲームの均衡解を求めるために，まず共同体メンバーが提案者の説得を受容するときの期待利得を求めると，以下のとおり。

$$\begin{aligned} \mathrm{E}U_m &= \beta_t\{h(g-c)+(1-h)(-c)\} \\ &\quad +(1-\beta_t)\{l(g-c)+(1-l)(-c)\} \\ &= \beta_t(hg-c)+(1-\beta_t)(lg-c) \\ &= \beta_t g(h-l)+lg-c \end{aligned}$$

ただし，β_t は，t 期に提案者の説得する行為が公共的なものである，という共同体メンバーの信念である。

メンバーが提案者の説得を拒否すると利得は０なので，$\mathrm{E}U_m > 0$ であれば，提案者の説得を受容する。上述のように，$h > l$ なので，$lg \geq c$ であれば常に $\mathrm{E}U_m > 0$ であり，$hg \leq c$ であれば常に $\mathrm{E}U_m > 0$ である。このモデル分析の意図は，公共的に利益になる「自己犠牲的」行為が，価値あるものとしてどのように定着していくかを分析することにあるので，常に受容されるケースや常に拒否されるケースは分析す

る必要が無い。従って，そうでないケース，即ち $lg < c$ かつ，$hg > c$ であるケースについて考える。

$\mathrm{E}U_m = \beta_t g(h-l)+lg-c > 0$ となるのは，$\beta_t > (c-lg)/g(h-l)$ のときであるが，$hg > c$ なので，この条件を満たす $0 \leq \beta \leq 1$ は存在する[15]。従って $\beta_t > (c-lg)/g(h-l)$ であれば，メンバーは提案者の説得を受容し，$\beta_t < (c-lg)/g(h-l)$ であれば拒否する。

このモデルの結果からは，それが専ら提案者の私的利益を意図したものか，公共的利益を意図したものかはわからない。しかし，もしその行動が公共的利益（g）を実現すれば，公共的行為であるという事後的信念はベイジアン・ルールに従って，β_t から

$$\beta_{t+1} = \beta_t h/\{\beta_t h+(1-\beta_t)l\}$$

に増加する。$(c-lg)/g(h-l)$ は所与なので，$t+1$ 期に提案者が再び説得すると，メンバーは再び受容する。なお，$\beta_0 = qx_s/\{qx_s+(1-q)x_p\}$ である。

逆に，もしその行動が公共的利益（g）を実現しなければ，公共的行為であるという事後的信念は，β_t から

$$\beta_{t+1} = \beta_t(1-h)/\{\beta_t(1-h)+(1-\beta_t)(1-l)\}$$

に減少する。$t+1$ 期に提案者が再び説得するとき，メンバーが再び受容するか拒否するかは，この減少の程度に依存する。このモデルでは，一度拒否されると β の値が増加するチャンスは無いので，そこでゲームは終了する。もしくは，次回以降，常に拒否される。

次に提案者が公共志向の提案をするときと，私的利益中心の提案をするときの期待利得を考える。公共志向の提案をするときの提案者の期待利得は，メンバーが提案者の説得を受容する行動戦略の確率を y とすると，

$$\begin{aligned}\mathrm{E}U_{PL} &= y\{h(b-c_L+g-c)+(1-h)\\ &\quad\times(-c_L-c)\}+(1-y)(-c_L)\\ &= y\{h(b+g)-c_L-c\}+(1-y)(-c_L)\\ &= y\{h(g+b)-c\}-c_L\end{aligned}$$

である。説得しないと利得は 0 なので，$\mathrm{E}U_{PL} > 0$ であれば，公共志向の提案者は説得を試みる。上式を変形すると $y > c_L/\{h(g+b)-c\}$。$0 \leq y \leq 1$ なので，$c_L/\{h(g+b)-c\} < 1$，即ち $h > (c+c_L)/(b+g)$ でなければ，提案者が説得を試みることはあり得ない。さらに，$0 \leq h \leq 1$ なので，$(c+c_L)/(b+g) < 1$，即ち $c+c_L < b+g$ でなければ，提案者が説得を試みることはあり得ない。

$h > (c+c_L)/(b+g)$ と $c+c_L < b_L+g$ が成立しているとき，提案者が，メンバーの行動戦略 y を，$y > c_L/\{h(g+b)-c\}$ と考えれば，提案者は説得を試みる。上述したように，初回でメンバーが拒否すれば，次回以降も常に拒否される。初回でメンバーが受容し，公共的利益（g）が実現し続ける限り，説得は常に受容されるが，公共的利益（g）が実現しないと，$\beta_t(1-h)/\{\beta_t(1-h)+(1-\beta_t)(1-l)\}$ が，$(c-lg)/g(h-l)$ より大きいかどうかで，次回に受容されるか拒否されるかが決まる。

私的利益中心の提案をするときの提案者の期待利得は，

$$\begin{aligned}\mathrm{E}U_{SL} &= y\{l(g-c_L^{\#}-c)+(1-l)\\ &\quad\times(b^{\#}-c_L^{\#}-c)\}+(1-y)(-c_L^{\#})\\ &= y\{l(g-b^{\#})+b^{\#}-c_L^{\#}-c\}+(1-y)(-c_L^{\#})\\ &= y\{l(g-b^{\#})+b^{\#}-c\}-c_L^{\#}\end{aligned}$$

である。説得しないと利得は 0 なので，$EU_{SL} > 0$ であれば，自己利益志向の提案者は説得を試みる。そのためには第 1 項が正でなければならないが，$y \geq 0$, $g < b^{\#}$ なので，そうなるのは $l(b^{\#}-g) < b^{\#}-c$ のとき。そのとき $\mathrm{E}U_{SL} > 0$ となる条件は，$y > c_L^{\#}/\{l(g-b^{\#})+b^{\#}-c\}$。$0 \leq y \leq 1$ なので，$c_L^{\#}/\{l(g-b^{\#})$

15) $(c-lg)/g(h-l) < 1$ であれば，条件式を満たす $0 \leq \beta \leq 1$ が存在する。不等号の両辺に左辺の分母を掛けると，$h-l > 0$ なので，$c-lg < gh-gl$，従って，$c < gh$。

$+b^{\#}-c\} < 1$，即ち $l(g-b^{\#}) > c+c_L{}^{\#}-b^{\#}$ でなければ，提案者が説得を試みることはあり得ない。$g < b^{\#}$ なので，この式は，$l < (b^{\#}-c-c_L{}^{\#})/(b^{\#}-g)$ と変形できる。この不等式の右辺は正なので，$l < \min[1, (b^{\#}-c-c_L{}^{\#})/(b^{\#}-g)]$ のとき，提案者は説得を試みる可能性がある。

このとき，提案者が，メンバーの行動戦略 y を，$y > c_L{}^{\#}/\{l(g-b^{\#})+b^{\#}-c\}$ と考えれば，提案者は説得を試みる。公共的提案と同様，初回でメンバーが拒否すれば，次回以降も常に拒否される。初回でメンバーが受容し，公共的利益（g）が実現し続ける限り，説得は常に受容されるが，公共的利益（g）が実現しないと，$\beta_t(1-h)/\{\beta_t(1-h)+(1-\beta_t)(1-l)\}$ が，$(c-lg)/g(h-l)$ より大きいかどうかで，次回に受容されるか拒否されるかが決まる。

最後に，この自己犠牲的行為が十分な公共的利益を生むことが繰り返し実績として証明されれば，前節の躾の時と同じような過程を経て，提案が公共的であるという信念が社会的に形成されていく。ただし，これはあくまでも合理的に考えて，この一見犠牲的な行為が実は高い利得を生むことを共同体メンバーが信じるようになっただけであり，この行為が倫理性を得たわけではない。

しかし，このような「公共的」行為が慣習化すれば，一々その行為を正当化する理屈を考えることは無くなり，行為自体を「善いこと」だと感じるようになる。即ち，「倫理規範」に転化する[16]。当然，規範に従うことに正の効用が伴い，規範からの逸脱には負の効用が伴う。また，他人の行動が規範から逸脱している場合も，それは社会に損失を与える（「悪い」）行為だと判断される。したがって，他人が規範から逸脱するのを防ぐ行為も規範に沿った行為であり，正の効用を生む。それが自己犠牲（負の効用）を伴う懲罰的行動だとしても，倫理的満足感の正の効用が上回れば合理的な選択である。このような過程を経てこの行為が倫理的行為だという価値観を持った人の数が，一定の数まで（限界質量を超える程度まで）増えれば，その倫理が社会的に維持できる好循環を生み出せることは，上述の Bowls and Gintis（2003）によって証明されている。こうして，大人の世界に新たな倫理規範が定着し，それが子供に躾けられるという循環が成立する。

16) この過程は子供の躾と同様効用関数の変化を伴うが，子供のように因果関係を理解していないのではなく，因果関係を忘れるのである。これは記憶の節約という意味で合理的とも言える。

5．考察：倫理規範の形成の現代的課題

本稿では倫理的規範の形成過程を，大人が子供に社会的ルールを躾ける内面化過程と，リーダーが共同体メンバーに，ある行為が社会的ルール足りうることを説得する過程に分けてモデル化した。前者では，とくに大人のインセンティブを考えることで，従来の学習モデルとは異なる躾のゲームを提示した。大人が既に倫理的規範を身につけていることを前提にすると，進化ゲームを使った倫理や規範の慣習化モデルを応用することができた。

後者は，情報の非対称性と学習ゲームの応用であるが，本稿では，異質なプレーヤの中から集合行為のイニシアティブを取るプレーヤ（リーダー）が出現することを仮定し，従来の動学的進化ゲームの前提となっていた倫理的互恵的動機発生のきっかけ及び生成過程を明示したモデルを示した。この二つの過程を組み合わせて，新たな規範形成・浸透過程の総合的モデルを提示した。

本稿のモデルのように，徐々に形成された倫理的規範は，経験や論理によって説得され続けなければ維持できない。例えば，リーダー（たち）の行為が余りに利己的であると思われれば，彼らが作った規範体系の説得力は失われていき，

反抗勢力の拡大，ひいては支配層交代ということも起こり得る。

リーダーたちが利己的でなくても，社会の変容に応じて公共性の意味が変化することもある。近代以降の人間社会で進行してきた，生産技術の向上（生産性の飛躍的上昇），エネルギー，輸送，情報技術の進歩は，一方で，安全や健康面での危険を減少させ，豊かで自由な生活を実現させてきた。しかし他方で，技術進歩や（その結果の）社会の巨大化が，所属する共同体と公共サービスを提供する共同体のズレを生じさせ，公共的義務に対する共同体からの心理的圧力を著しく退化させ，モラルハザードやフリーライダー（あるいは社会的ジレンマ）の問題を拡大させた。現在の日本で倫理的規範の崩壊が懸念されているのは，この後者の問題であろう。

本稿のモデルに即して言えば，この問題には二つの側面がある。一つは，期待利得（g），費用（c_L），成功確率（h）などのパラメータが変化したと考えられる。パラメータの値は，上述のような社会環境や政治システムなどの制度によって規定されるが，それは歴史的に形成されてきたものである。短期的には安定していると仮定できるが，戦後の社会的制度の制度疲労が90年代に表面化した結果，今まさにこうしたパラメータが大きく変化していると考えられる[17]。

しかし，パラメータの値が結局は歴史的に形成されてきたものであれば，しかもそれが国のような大きな社会のパラメータであれば，それを人為的に操作することは難しい。大きな社会の管理運営は官僚組織や政党などの行政的政治的専門集団の運営に任されており，それをマスコミが監視したり，利益団体が特定領域で政治的影響力を行使したりするのみである。一般個人の影響力は限られており，したがって社会に対する関心も希薄になるのは当然である。その結果，必要な公共財・サービスがそれなりに供給されていて，その費用＝税金がそれなりの水準に抑えられていれば大きな不満は表出しない。個人が主体的に政治的行動に出るのは，自分の被るデメリットがある閾値を越えて，しかもそれに対する不満を政治的行動に移すことによって得られる期待利得（不満解消の心理的満足感も含めて）がその政治的行動の費用を上回るときである。大きな社会であればその閾値はきわめて高いと推測される。この観点からすると，新たな倫理規範形成のきっかけとなるような環境変化を人為的に作るには，政治主導の制度改革や地域や分野を特定した実験的改革が現実的な選択肢であろう。

本稿のモデルから分析できるもう一つの側面は，第一の側面の環境変化に対応して，個人の倫理規範がどのように修正され維持されていくのかという問題である。現代の個人は，巨大社会の一員であると同時に，依然として狭い仲間社会の一員でもある。社会化という点では，家族，地域や職場，親しい友人のグループという仲間内からの影響の方が依然として大きい。そして仲間内のルール（倫理規範）は，それが非公式なだけに，むしろ公式的な「社会」のルールより，歴史的な倫理観との継続性は高い。ただしこの次元での社会化は，本稿のモデルの前半，つまり躾の部分に相当し，後半の「社会」的倫理規範の修正・維持がうまく機能していなければ，規範維持の循環が断ち切られる。公式には戦前の倫理規範が否定された後，新たな倫理規範が体系的に形成されなかった結果，子供が反倫理的行為をとったときに，大人がペナルティを科すことの倫理的価値を裏付ける論理も弱く，そのことの倫理的価値が次第に低下しているのではないだろうか。したがって，本稿のモデルにおけるペナルティ（p）や躾の定着率（ρ）といったパラメータの値が低下し，倫理規範維持に不利になったと考えられる。

最近の日本における倫理規範喪失への懸念の

17）この過程は制度変化のゲームモデルにおける断続平衡（punctuated equilibrium）（前掲 Aoki）の一つの具体例とも言える。

高まりは，倫理的規範形成・維持のメカニズムに問題が生じているという認識が社会の中で共有されつつあることの現れである。しかし，現代の日本に適合的な新たな倫理規範として有力な提案があるわけではない。逆説的ではあるが，今求められているのは，本稿の初めに言及した社会的選択論の「何が倫理的か」という問いに対する答えであるようにも見える。とくに本稿のモデルでは，全く倫理的規範の無い時点からのモデル化を考えたため，当初の社会的ルールの提案に何の倫理性も無かった。したがってこの時点で公理的に倫理性を考えるべきだという考え方は，却って受け入れ易いかもしれない[18]。しかし，いくら倫理的規範が崩壊しているといっても，現実のリーダーの提案には，彼（女）の個人的歴史や育った共同体に特有の倫理的価値観が含まれざるを得ない。それと同様，潜在的な新しい倫理規範は，日々倫理的問題に取り組んでいる教育現場や福祉現場といった現在の共同体・社会における実践的な取組みの中で発見され提案されているはずである。個々の共同体が新たな（あるいは修正された）共同体倫理（ルール）を発見し浸透させ，それが社会全体に広がっていくというメカニズムが本稿のモデルから演繹される倫理規範形成過程の姿である[19]。

18）この初期時点は Rawls（1971）の veil of ignorance の original position と似た概念と解釈できよう。

19）同じニュアンスを持つ提案として，Binmore（1994, 98）は，上記 Rawls の original position の代わりに status quo を規範形成の出発点として考えることを提唱している。

参考文献

［1］ Aoki, Masahiko（2001）“The Subjective Game Form and Institutional Evolution as Punctuated Equilibrium”in *Towards a Comparative Institutional Analysis*, MIT Press.（滝澤・谷口訳（2001）『比較制度分析に向けて』NTT 出版）

［2］ Axelrod, Robert（1984）*The Evolution of Cooperation*, Basic Books.

［3］ Axtel, Robert, Epstein, Josua, and Young, Peyton（2001）“The Emergence of Classes in a Multi-Agent Bargainning Model,”in Durlauf and Young ed. *Social Dynamics*, MIT Press.

［4］ Binmore, Ken（1994, 98）*Game Theory and Social Contracts I, II*, MIT Press.

［5］ Bowls Samuel and Gintis, Herbert（2003）“The Evolution of Strong Reciprocity：Cooperation in Heterogeneous Populations,”*Theoretical Population Biology*, forthcoming.

［6］ Falk, Armin and Fishbacher, Urs（2000）“A Theory of Reciprocity,”Working Paper No.6, July 2000, Institute for Empirical Research in Economics, University of Zurich.

［7］ Fehr, Ernst and Gachter, Simon（2000）, “Cooperation and Punishment in Public Goods Experiment,”*American Economic Review*, Vol. 90 (4), pp.980-994.

［8］ Fudenberg, Drew and Levine, David K.（1998）*The Theory of Learning In Games*, MIT Press.

［9］ Kandori, Michihiro（2003）“The Erosion and Sustainability of Norms and Morale,”*The Japanese Economic Review*, Vol.54 (1), pp.29-48.

[10] Kandori, Michihiro, Mailath, George J., and Rob, Rafael（1993）“Learning, Mutation, and Long-run Equilibria in Games,”*Econometrica*, Vol. 61, pp.29-56.

[11] Kaneko, Mamoru and Matsui, Akihiko

(1997)“Inductive Game Theory：Discrimination and Prejudice,”*Journal of Public Economic Theory*, Vol. 1 (1), pp.101-137.

[12] 町野和夫 (2003)「倫理的規範形成の二段階モデル」Discussion Paper Series B No.41，北海道大学経済学研究科.

[13] 無藤隆・久保ゆかり・遠藤利彦 (1995)『発達心理学』岩波書店.

[14] Rawls, John (1971) *A Theory of Justice,* Harvard University Press.（矢島鈞次 監訳 (1979)『正義論』紀伊国屋書店）

[15] 斎藤耕二，菊池章夫編 (1990)『社会化の心理学ハンドブック：人間形成と社会と文化』川島書店.

[16] 作田啓一 (1972)『価値の社会学』岩波書店.

[17] Simon, Herbert A. (1956),“Rational Choice and the Structure of the Environment,”*Psychological Review*, Vol. 63, pp.129-138.

[18] 鈴村興太郎 (2000)「厚生経済学の情報的基礎」岡田・神谷・黒田・伴編『現代経済学の潮流 2000』第１章，東洋経済新報社.

[19] Young, Peyton, H. (1993)“The Evolution of Convention,”*Econometrica*, Vol. 61, pp.57-84.